Erik Ter-Gabrielyan
STAT100 Section 0122

Problem Set 4

1a.
The explanatory variable would be the Server.

1b.
The response variable would be the PctTip, or Percent Tip aka how much of the bill is tipped to the server.

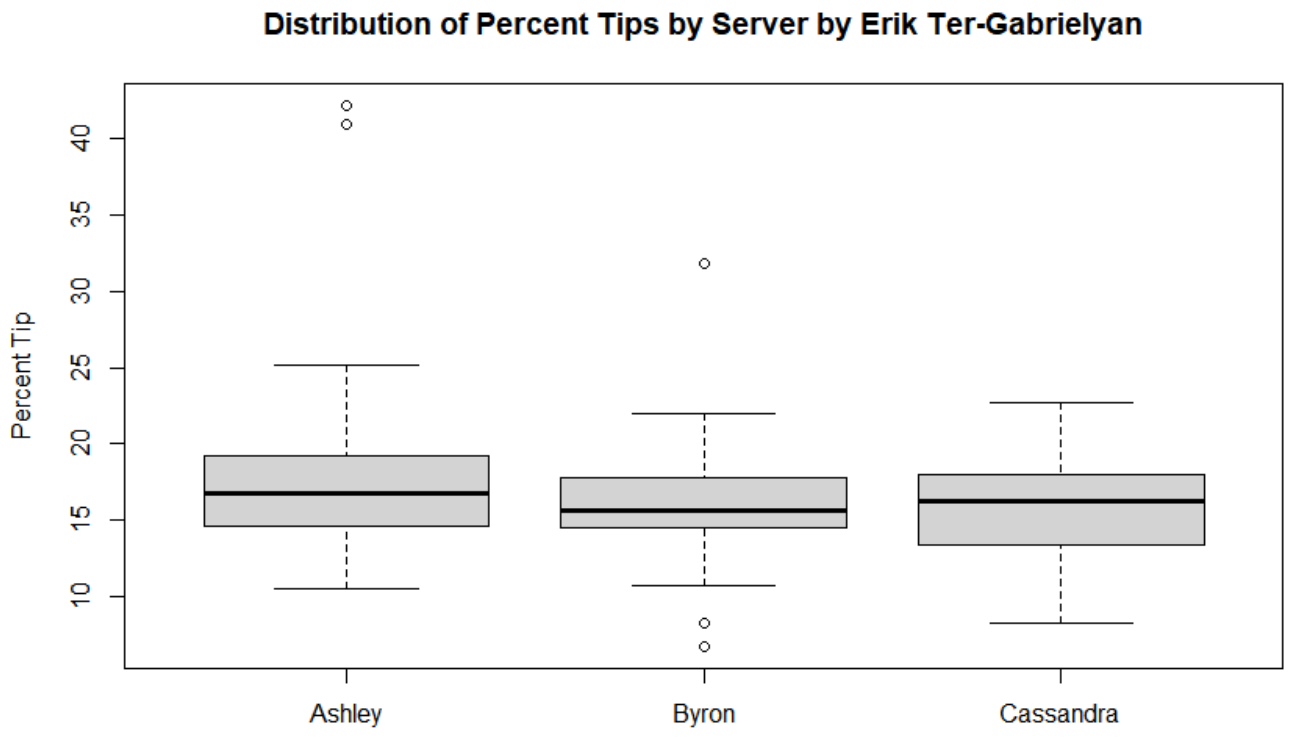
1c.
The role-type classification for this would be C->Q.

1d.

```
> # Erik Ter-Gabrielyan
> tapply(Restaurant_Tips$PctTip,Restaurant_Tips$Server,summary)
$Ashley
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  10.50   14.65   16.75   17.54   19.23   42.20

$Byron
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   6.70   14.50   15.60   16.02   17.80   31.80

$Cassandra
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   8.20   13.38   16.25   16.11   18.00   22.70
> boxplot(Restaurant_Tips$PctTip~Restaurant_Tips$Server,main="Distribution of Percent Tips by Server by Erik Ter-Gabrielyan",xlab="Server",ylab="Percent Tip",)
```



1e.
Ashley had the highest median tip percentage with 16.75%.

1f.
Byron had the lowest median tip percentage with 15.60%.

1g.
I would say that there is not a significant enough difference in the tip percentage between the different servers. The difference between the highest and lowest median tip percentage is only 1.25%, the difference between the highest and lowest mean tip percentage is only 1.52%, and Ashley's data is skewed with two massive tips over 40%. It is not possible to tell if these are because of the quality of Ashley's service or simply random chance of landing a generous customer at her table.

2a.
The explanatory variable would be the Day.

2b.
The response variable would be the Server.

2c.
The role-type classification would be C->C.

2d.

```
> # Erik Ter-Gabrielyan
> table(Restaurant_Tips$Day,Restaurant_Tips$Server)

     Ashley Byron Cassandra
  Fri     10    10         6
  Mon      0    20         0
  Thu     13    14         9
  Tue      9     0         4
  Wed     28    21        13

> table1 = table(Restaurant_Tips$Day,Restaurant_Tips$Server)
> 100*table1/rowSums(table1)
         Ashley     Byron Cassandra
  Fri  38.46154  38.46154  23.07692
  Mon   0.00000 100.00000   0.00000
  Thu  36.11111  38.88889  25.00000
  Tue  69.23077   0.00000  30.76923
  Wed  45.16129  33.87097  20.96774
```

2e.
Ashley served 38.46% of the tables, Byron served 38.46% of the tables, and Cassandra served 23.08% of the tables on Friday.

2f.
On Monday, neither Cassandra or Ashley served any tables, only Byron. Also, on Tuesday, Byron did not serve any tables, only Cassandra and Ashley did.

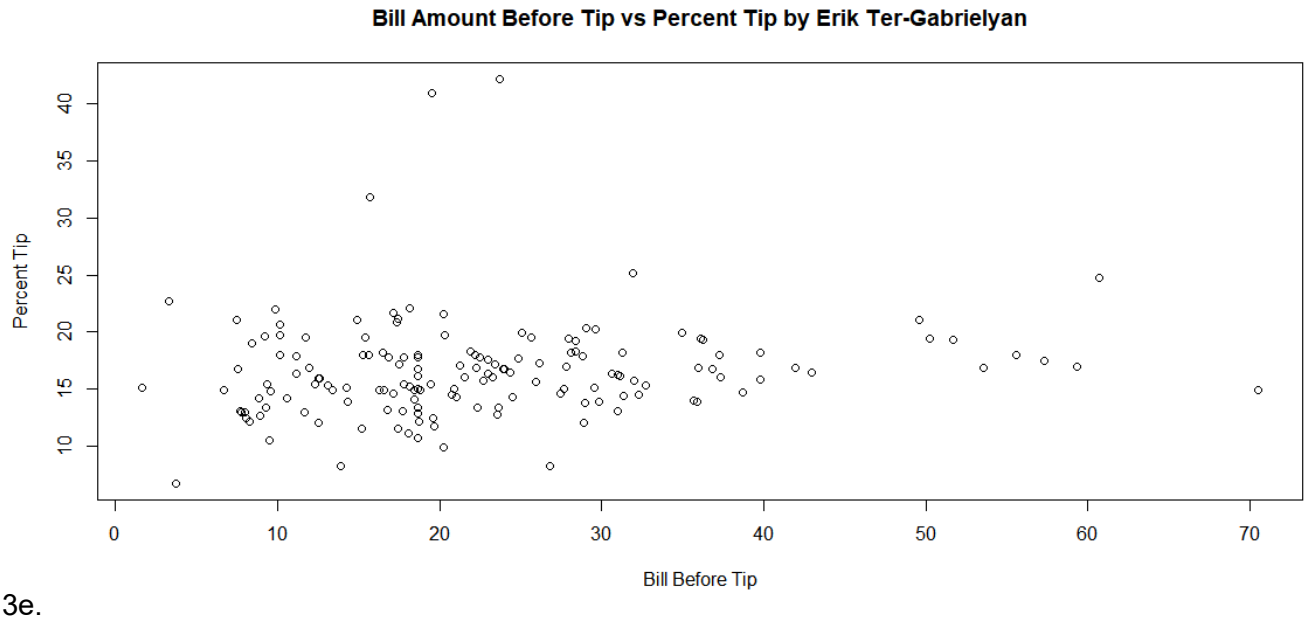
2g.
There certainly is a relationship between the day and server. On certain days, only certain servers worked, and certain servers served way more tables than others on certain days, rather than it being a 33.33/33.33/33.33 split between all three of the servers.

3a.
The explanatory variable would be the Bill.

3b.
The response variable would be the Percent Tip.

3c.
The role-type classification is Q->Q.

3d.
> # Erik Ter-Gabrielyan
> plot(Restaurant_Tips$Bill,Restaurant_Tips$PctTip,xlab = "Bill Before Tip",ylab = "Percent Tip",main = "Bill Amount Before Tip vs Percent Tip by Erik Ter-Gabrielyan")



3e.
> cor(Restaurant_Tips$Bill,Restaurant_Tips$PctTip)
[1] 0.1352976

3f.
The correlation coefficient shows it is a positive relationship, but also very close to having no relationship.

3g.
Bill = 16.497 − 0.375(Bill Before Tip)

3h.
50 = 16.497 - 0.375(Percent Tip)
33.503 = -0.375(Bill Before Tip)

3i.
Since the correlation coefficient is so close to 0, there is not a significant difference ie. you cannot predict percent tip by bill.